ARO 23699.22-MA

2

AD-A213 741

Nonparametric estimation of trends in linear stochastic systems

Ian W. McKeague*

Department of Statistics
Florida State University, Tallahassee, Florida 32306-3033

and

Tiziano Tofoni

Postgraduate School of Telecommunications
G. Reiss Romoli, 67100 L'Aquila, Italy

FSU Technical Report No. M-814
USARO Technical Report No. 107

September, 1989

AMS 1980 subject classifications. 62G05, 62M09, 93E12.

Key words and phrases. Kalman-Bucy filtering, linear periodic systems, nonparametric trend estimation, method of sieves, kernel estimators, Volterra integral equations.

---

* Research supported by Army Research Office Grant DAAL03-86-K-0094.

Abstract

Techniques for the estimation of unknown additive trends present in the state and measurement processes of a Kalman–Bucy linear system are introduced. We obtain asymptotic results describing the performance of the estimators under i.i.d. and periodic observation schemes. The observed process is given by $dY(t) = g(t)\,dt + dZ(t)$, where Z is the measurement process and g is an unknown trend function, and there is an additive trend f present in the state process X. These two cases need to be treated separately in order to ensure identifiability. The problem is to estimate f and g, and remove them from the measurement process. Trend removal involves replacing f and g in the Kalman filter $\hat{X}(t) = E(X(t)|\mathcal{F}_t^Y)$—based on observation of Y—by appropriate estimates. We show that this can be done under the following observation schemes: (I) n i.i.d. replicates of Y over a fixed interval $[0, T]$, (II) observation of a single trajectory of Y over a long interval $[0, nT]$, where f, g and the functions defining the linear system are periodic with period T.

1. Introduction

Consider a linear stochastic system of the type introduced by Kalman and Bucy (1961): A p-dimensional "state" process X and a q-dimensional "measurement" process Z are given by the stochastic differential equations

$$dX(t) = A(t)\,X(t)\,dt + B(t)\,u(t)\,dt + dW(t)$$
$$dZ(t) = C(t)\,X(t)\,dt + dV(t)$$

$0 \le t \le T$, where W and V are independent p and q-dimensional Wiener processes, $u(\cdot)$ is a known deterministic input, A, B, C are known non-random time-varying matrices of suitable dimensions, $X(0)$ is independent of W and V, the mean $E(X(0)) = m$ and covariance matrix of $X(0)$ are known, and $Z(0) = 0$. The Kalman filtering theory provides recursive formulae for the conditional expectation $\hat{X}(t) = E(X(t)|\mathcal{F}_t^Z)$ which is the optimal mean square estimate of the state $X(t)$ given the past $\mathcal{F}_t^Z = \sigma(Z_s, 0 \le s \le t)$ of the measurement process, see Liptser and Shiryayev (1978) and Kallianpur (1980).

In real applications of the Kalman filter to signal processing it is often found that unknown additive trends are present in the state and measurement processes; that is, the state process X is given by

$$dX(t) = f(t)\,dt + A(t)\,X(t)\,dt + B(t)\,u(t)\,dt + dW(t) \tag{1}$$

and instead of observing Z, we observe the process Y given by

$$dY(t) = g(t)\,dt + dZ(t), \qquad Y(0) = 0, \tag{2}$$

where f and g are unknown "trend" functions.

In the present paper we shall consider the problem of estimating the trends f and g and removing them from the measurement process. Trend removal amounts to replacing the functions f and g used in the Kalman filter $\hat{X}(t) = E(X(t)|\mathcal{F}_t^Y)$—based on observation of Y—by appropriate estimates $\hat{f}$ and $\hat{g}$.

Two types of observation scheme are considered:

(I) n realizations $\{Y_i(t),\ t \in [0,T],\ i = 1,\dots,n\}$ of the process Y satisfying (1) and (2) with the corresponding system realizations having independent noise processes W_i and V_i, $i = 1,\dots,n$.

(II) observation of a single trajectory of Y over the interval $[0, nT]$, where the functions f, g, A, B and C are periodic with period T.

Observation scheme (II) is relevant to situations where there is a "time-of-day" or "seasonal" effect present in the model; for example, in the analysis of circadian rhythm data in biology, or in the study of cyclic systems in control engineering—see the review article of Bittanti and Guardabassi (1986). We are interested in the asymptotic properties of estimators of f and g as $n \to \infty$ with T remaining fixed. We shall see that f and g are not identifiable unless one of them is absent from the model (i.e. $f = 0,\ g \neq 0$ or $f \neq 0,\ g = 0$).

There is a vast literature on the estimation of finite dimensional parameters in discrete time linear stochastic systems; refer to the books of Davis and Vinter (1984) and Kumar and Varaiya (1986). In continuous time such problems were first studied by Balakrishnan (1973). Further contributions have been made by Bagchi (1980), Tugnait (1980) and Bagchi and Borkar (1984). Nonparametric estimation for linear stochastic systems is considered to be a difficult problem; see, for instance, the closing comment of a recent paper of Aihara and Bagchi (1989). In general the functions A, C, f and g are not even identifiable. In the present paper we are studying the very special case in which A, B and C are known, and at least one of the trend functions is known to be absent.

There is an extensive literature on nonparametric estimation for the drift (or trend) function, g, in a diffusion process satisfying (2) with Z as a Wiener process; see Ibragimov and Khasminski (1980, 1981), Geman and Hwang (1983), Nguyen and Pham (1982), Beder (1987), McKeague (1986)—who allowed Z to be a general square integrable martingale, and Leskow (1989)—who considered the case of a periodic model. These authors use either Parzen-Rosenblatt type kernel estimators or Grenander (1980) sieve estimators for g, but those estimators are not directly applicable to the present setting, unless C is identically zero (in which case only g is identifiable). We shall find that there is a function h, related to g and f through two Volterra integral equations, and h can be estimated by kernel or sieve type estimators. Estimates of g and f can then be obtained by inserting estimates of h or its first derivative h' in the solutions of the Volterra integral equations.

The paper is organized as follows. Section 2 contains introductory discussion concerning the basic innovations representation of the observation process, identifiability, bias under misspecified trends, and schemes (I) and (II). Estimation of the trend in the measurement process under schemes (I) and (II) is treated in Sections 3 and 4 respectively. In Section 5 we consider estimation of the trend in the state process. In these sections, to simplify the presentation, we assume that the state and measurement processes are one-dimensional ($p = q = 1$). Section 6 contains remarks on the multi-dimensional case. In Section 7 we indicate some directions for further work.

To conclude this section we shall briefly put our problem in perspective with other inference problems for stochastic processes. Statistical models for stochastic processes are of two broad types. If we observe a process $Y = (Y_t, t \geq 0)$ and we have a covariate process $X = (X_t, t \geq 0)$ to incorporate into the analysis, then we may consider a *partially specified* model in which, loosely speaking (see Greenwood (1988) for a more precise definition), only the conditional distribution of Y given X is specified in terms of an unknown parameter θ. Alternatively, we may know the full joint distribution of (Y, X) for each θ, in which case we have a *fully specified* model. Partially specified models are especially useful and widely applied in the analysis of life history data by taking Y as a counting process describing the times of events in the life of an individual, and X representing a covariate process specific to the individual—the structure of the marginal distribution of X being unspecified; see Arjas and Haara (1984) and Andersen et al. (1988). Fully specified models on the other hand are widely used in the engineering sciences where precise models for the covariate process X can often be developed from well-understood system dynamics. Our model is of this latter type.

Observation schemes may similarly be classified into two broad types: partial and full. In the survival analysis setting partial observation may arise from censoring, truncation, or grouping of the data, see Andersen et al. (1988) and McKeague (1988). It arises in the stochastic systems setting when the state of the system is observed in the presence of noise, as in (1). Despite the diverse applications of such schemes and models, there is a surprising unity to the techniques used. For example, our kernel function techniques are similar to the methods used by Ramlau-Hansen (1983) for the estimation of counting process intensities, and our approach to the periodic case in some ways resembles that of Pons and de Turckheim (1988) to Cox's periodic regression model.

2. The innovations representation

We shall assume throughout that the functions f, g, A, B, C and u are smooth, and $C(t)$ does not vanish anywhere on $[0, T]$. The equations for the Kalman-Bucy filter (see Kallianpur 1980, Section 10.3) are

$$\begin{aligned} d\hat{X}(t) &= [f(t) + A(t)\hat{X}(t) + B(t)\,u(t)]\,dt + D(t)\,d\nu(t) \\ d\nu(t) &= dY(t) - [g(t) + C(t)\hat{X}(t)]\,dt, \end{aligned} \tag{3}$$

where $\hat{X}(0) = m$. The process ν is the so-called *innovations process* which is known to be a standard Wiener process. The function D is the *Kalman gain* which in the present set-up does not depend on f or g. In fact $D(t) = C(t)P(t)$, where P is the unique positive solution to the *Riccati differential equation*

$$P'(t) = 2A(t)P(t) - C^2(t)P^2(t) + 1. \tag{4}$$

with initial condition $P(0) = \mathrm{Var}(X(0))$. From (3) we have

$$d\hat{X}(t) = [A(t) - D(t)\,C(t)]\hat{X}(t)\,dt + [f(t) + B(t)\,u(t) - D(t)\,g(t)]dt + D(t)\,dY(t).$$

Using Theorem 4.2.4 of Davis (1977) we can solve this equation for $\hat{X}$. Substituting the solution into the the second equation in (3) we obtain the following *innovations* *representation* for Y:

$$Y(t) = \int_0^t [h(s) + U(s)]\, ds + \nu(t), \tag{5}$$

where

$$h(t) = g(t) + C(t) \int_0^t \Psi(t,s)[f(s) - D(s)\, g(s)]\, ds. \tag{6}$$

Here $\Psi(t,s)$ is the solution to the linear time-varying system

$$\frac{\partial \Psi(t,s)}{\partial t} = [A(t) - D(t)\, C(t)]\Psi(t,s), \qquad \Psi(s,s) = 1$$

and U is given by

$$U(t) = C(t)\Big\{\Psi(t,0)\, m + \int_0^t \Psi(t,s)[B(s)\, u(s)\, ds + D(s)\, dY(s)]\Big\}.$$

The representation (5) will be of prime importance in the sequel.

Identifiability of f and g.

We see from (5) that the function h is identifiable given observation of Y and U; however, f and g are identifiable only in so far as they are uniquely determined in terms of h through (6). Thus, the functions f and g are not in general simultaneously identifiable from observation of Y. However, if the trend is absent from the measurement process ($g = 0$) then (6) reduces to

$$h(t) = \int_0^t \Phi(t,s)\, f(s)\, ds, \tag{7}$$

where $\Phi(t,s) = C(t)\, \Psi(t,s)$. If the trend is absent from the state process ($f = 0$) then (6) reduces to

$$h(t) = g(t) + \int_0^t \Gamma(t,s)\, g(s)\, ds. \tag{8}$$

where $\Gamma(t,s) = -\, C(t)\, \Psi(t,s)\, D(s)$.

As equations involving the unknown f and g, (7) and (8) are *linear Volterra integral equations* of the *first* and *second* kind respectively. It follows from standard results on Volterra equations (see Linz, 1985) that (8) has a unique solution for g, and (7) has a unique solution for f provided $C(t)$ does not vanish on $[0,T]$. Since h is identifiable, the trend f is identifiable when $g = 0$, and g is identifiable when $f = 0$.

The log-likelihood function.

The innovations representation (5) allows us to write down an explicit expression for the log-likelihood function $L(h) = \log[d\mu_h/d\mu_W](Y)$, where μ_h is the measure induced on $C[0,T]$ by Y, and μ_W is Wiener measure. By Liptser and Shiryayev (1977, Theorem 7.7) we have that $\mu_h \ll \mu_W$ and

$$L(h) = \int_0^T \pi(s)\,dY(s) - \frac{1}{2}\int_0^T \pi^2(s)\,ds \tag{9}$$

where $\pi(s)$ is the term inside the square brackets in (5).

The bias caused by misspecified trends.

What is the effect on the mean square error of the Kalman filter (3) of using incorrect trend functions $f^* \neq f$, $g^* \neq g$? The answer to this question should provide us with a *modus operandi* for choosing estimators $\hat{f}$, $\hat{g}$ to be used in place of the unknown f, g. Let $\hat{X}^{f,g}(t)$ denote the Kalman filter estimate of $X(t)$ based on (3). The *bias* caused by using f^*, g^* instead of f, g at time t,

$$BIAS(f^*, g^*, t) \equiv \hat{X}^{f^*,g^*}(t) - \hat{X}^{f,g}(t),$$

can be found from (3), cf. Jazwinski (1970, p.252),

$$BIAS(f^*, g^*, t) = \int_0^t \Psi(t,s)[f^*(s) - f(s) + D(s)\,(g(s) - g^*(s))]\,ds.$$

The increase in the mean square error caused by using estimators f^*, g^* instead of f, g is solely due to this nonrandom bias and is given by $[BIAS(f^*, g^*, t)]^2$.

Observation scheme (I).

The processes associated with the ith realization are given the subscript i, as in ν_i, u_i, U_i, $L_i(h)$ etc.. Note that although the observed processes $\{Y_i,\ i = 1, \ldots, n\}$ are independent, they are not necessarily identically distributed since the inputs u_i are not assumed to be identical for each i. However, the innovations processes ν_i are i.i.d. Wiener processes. From (5) we have

$$Y_i(t) = \int_0^t [h(s) + U_i(s)]\,ds + \nu_i(t), \tag{10}$$

where

$$U_i(t) = C(t)\Big\{\Phi(t,0)\,m + \int_0^t \Phi(t,s)[B(s)\,u_i(s)\,ds + D(s)\,dY_i(s)]\Big\}.$$

The log-likelihood function $L^{(n)}(h)$ is given by

$$L^{(n)}(h) = \sum_{i=1}^{n} L_i(h).$$

Observation scheme (II).

Scheme (II) can be treated using a similar framework to scheme (I). Let h_i, U_i, Y_i and ν_i be the following restrictions of h, U, Y and ν to the i-th period:

$$\begin{aligned} h_i(t) &= h(iT+t) \\ U_i(t) &= U(iT+t) \\ Y_i(t) &= Y(iT+t) - Y(iT) \\ \nu_i(t) &= \nu(iT+t) - \nu(iT) \end{aligned}$$

$0 \le t \le T$. These processes satisfy

$$Y_i(t) = \int_0^t [h_i(s) + U_i(s)]\, ds + \nu_i(t). \tag{10'}$$

Since ν is a Wiener process (which has stationary independent increments), the processes ν_i, $i = 1, \dots, n$ are also i.i.d. Wiener processes on $[0, T]$. The log-likelihood function $L^{(n)}(h)$, given by (9) with with T replaced by nT, can be written as

$$L^{(n)}(h) = \sum_{i=1}^{n} L_i(h_i).$$

Note that the function h is not periodic so that $h_i \neq h$. This is the basic difference between schemes (I) and (II), making (II) much harder to analyze.

3. Trend in the measurement process—i.i.d. case

In this section we consider estimation of g under scheme (I) with $f \equiv 0$. First we introduce estimators $\hat{g}$ of g such that $BIAS(0, \hat{g}, t) \to 0$ uniformly in t as $n \to \infty$ a.s.. In fact, *a fortiori*, $\hat{g}$ will be shown to be strongly L^2-consistent in the sense that $\|g - \hat{g}\| \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$, where $\|\cdot\|$ denotes the norm in $L^2[0, T]$.

The basic idea is to take as an estimator of g the solution $\hat{g}$ of the Volterra integral equation

$$\hat{h}(t) = \hat{g}(t) + \int_0^t \Gamma(t, s)\, \hat{g}(s)\, ds, \tag{11}$$

where $\hat{h}$ is an estimator of h. Note that the estimator so obtained is well defined since (11) admits a unique solution whenever $\hat{h} \in L^2[0, T]$—see Davis (1977, p. 125). Moreover, should $\hat{h}$ be a strongly L^2-consistent estimator of h, the following theorem shows that $\hat{g}$ is also strongly L^2-consistent.

THEOREM 1. *Let $\hat{h}$ be a strongly L^2-consistent estimator of h. Then the solution of the Volterra integral equation (11) is a strongly L^2-consistent estimator of g.*

PROOF: Let $\bar{h} = \hat{h} - h$ and $\bar{g} = \hat{g} - g$ be the estimation errors of $\hat{h}$ and $\hat{g}$ respectively and denote $M = \sup_{t,s\in[0,T]} |\Gamma(t,s)| < \infty$. Now from (8) and (11)

$$\bar{g}(t) = \bar{h}(t) - \int_0^t \Gamma(t,s)\,\bar{g}(s)\,ds$$

so that

$$|\bar{g}(t)|^2 \leq 2TM^2 \int_0^t |\bar{g}(s)|^2\,ds + 2|\bar{h}(t)|^2.$$

Using Gronwall's inequality (see Kallianpur, 1980, p.94) we then have

$$\begin{aligned}|\bar{g}(t)|^2 &\leq 4TM^2 \int_0^t |\bar{h}(s)|^2 \exp[2TM^2(t-s)]\,ds + 2|\bar{h}(t)|^2 \\ &\leq 4TM^2 \exp[2T^2M^2] \int_0^t |\bar{h}(s)|^2\,ds + 2|\bar{h}(t)|^2.\end{aligned}$$

Integrating this last inequality over the interval $[0,T]$ we easily get

$$\|\bar{g}\|^2 \leq (2 + 4T^2M^2\exp[2T^2M^2])\|\bar{h}\|^2.$$

This completes the proof. □

Orthogonal series sieve estimators for h.

The maximum of the log-likelihood function $L^{(n)}(h)$ is not attained when we maximize over the whole parameter space $L^2[0,T]$. The problem is that the parameter space is too large for the existence of the unconstrained maximum likelihood estimator. One remedy is to apply the *method of sieves* which consists in maximizing the log-likelihood function over an increasing sequence of subsets S_n, $n = 1, 2, \dots$ of the parameter space. We shall use an orthogonal series sieve $S_n = \text{span}\{\psi_r,\ r = 1, \dots, d_n\}$, where $\{\psi_r,\ r \geq 1\}$ is a complete orthonormal sequence in $L^2[0,T]$ and $d_n \to \infty$ as $n \to \infty$.

Let the coordinates of $h \in L^2[0,T]$ with respect to the basis $\{\psi_r,\ r \geq 1\}$ be denoted $(h_r,\ r \geq 1)$ and denote the vector $(h_1, \dots, h_{d_n})'$ by $\mathbf{h}^{(n)}$. Then, omitting terms not involving h, for $h \in S_n$

$$L^{(n)}(h) = \mathbf{h}^{(n)\prime}(\mathbf{Q}^{(n)} - \mathbf{P}^{(n)}) - \frac{n}{2}\mathbf{h}^{(n)\prime}\mathbf{h}^{(n)} \tag{12}$$

where $\mathbf{Q}^{(n)}$ and $\mathbf{P}^{(n)}$ are $d_n \times 1$ vectors with components

$$\mathbf{Q}_r^{(n)} = \sum_{i=1}^{n} \int_0^T \psi_r(t)\,dY_i(t)$$

$$\mathbf{P}_r^{(n)} = \sum_{i=1}^{n} \int_0^T \psi_r(t)\,U_i(t)\,dt.$$

Maximizing (12) with respect to $\mathbf{h}^{(n)}$ we obtain

$$\hat{h}(t) = \sum_{r=1}^{d_n} \hat{h}_r \, \psi_r(t) \tag{13}$$

where $\hat{\mathbf{h}}^{(n)} = [\hat{h}_1, \ldots, \hat{h}_{d_n}]'$ is given by

$$\hat{\mathbf{h}}^{(n)} = \frac{1}{n}(\mathbf{Q}^{(n)} - \mathbf{P}^{(n)}). \tag{14}$$

THEOREM 2. *Suppose that $d_n \to \infty$ and $d_n/n \to 0$ as $n \to \infty$. Then the orthogonal series sieve estimator $\hat{h}$ given by (13) is a strongly L^2-consistent estimator of h.*

PROOF: It suffices to show that $\|\hat{\mathbf{h}}^{(n)} - \mathbf{h}^{(n)}\| \xrightarrow{\text{a.s.}} 0$, where $\|\cdot\|$ can also denote the euclidean norm, depending on the context. By (10) and (14) the rth component of $\hat{\mathbf{h}}^{(n)} - \mathbf{h}^{(n)}$ is given by

$$(\hat{\mathbf{h}}^{(n)} - \mathbf{h}^{(n)})_r = n^{-\frac{1}{2}} \epsilon_r^{(n)}$$

where

$$\epsilon_r^{(n)} = n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^T \psi_r(t)\, d\nu_i(t).$$

for $r = 1, \ldots, d_n$. Thus

$$\|\hat{\mathbf{h}}^{(n)} - \mathbf{h}^{(n)}\|^2 = \frac{1}{n} \sum_{r=1}^{d_n} (\epsilon_r^{(n)})^2.$$

Now $\epsilon_r^{(n)}$, $r = 1, \ldots, d_n$ are i.i.d. $N(0,1)$ r.v.'s so that $n\|\hat{\mathbf{h}}^{(n)} - \mathbf{h}^{(n)}\|^2$ has a χ^2 distribution with d_n degrees of freedom. The proof is now completed using the Borel-Cantelli type argument given by Beder (1987, Section 5). □

Remark. The rate $d_n = o(n)$ is the best possible for L^2-consistency of the orthogonal series sieve estimators, cf. McKeague (1986) and Beder (1987).

Kernel estimators for h.

Let K be a bounded kernel function having integral 1, support $[-1, 1]$ and let $b_n > 0$ be a bandwidth parameter. Define

$$\tilde{h}(t) = \frac{1}{b_n} \int_0^T K\left(\frac{t-s}{b_n}\right) d\tilde{H}(s), \tag{15}$$

$$\tilde{H}(t) = \frac{1}{n} \sum_{i=1}^{n} \left\{ Y_i(t) - \int_0^t U_i(s)\, ds \right\}. \tag{16}$$

Here $\tilde{H}(t)$ estimates the function $H(t) = \int_0^t h(s)\,ds$.

THEOREM 3. *Suppose that $b_n \to 0$ and $b_n n^{1-\delta} \to \infty$ for some $0 < \delta < 1$. Then the kernel estimator $\tilde{h}$ given by (15) is a strongly L^2-consistent estimator of h.*

PROOF: First note that since h is continuous, $\|h^{(n)} - h\| \to 0$ where $h^{(n)}$ is the following smoothed version of h

$$h^{(n)}(t) = \frac{1}{b_n} \int_0^T K\left(\frac{t-s}{b_n}\right) h(s)\,ds.$$

It remains to show that $\|\tilde{h} - h^{(n)}\| \xrightarrow{\text{a.s.}} 0$ From (15)

$$\tilde{h}(t) - h^{(n)}(t) = (nb_n)^{-\frac{1}{2}} \epsilon^{(n)}(t),$$

where

$$\epsilon^{(n)}(t) = \frac{1}{\sqrt{b_n}} \int_0^T K\left(\frac{t-s}{b_n}\right) dW^{(n)}(s)$$

and $W^{(n)} = \sqrt{n}(\tilde{H} - H)$. It follows from (10) and (16) that $W^{(n)}$ is a standard Wiener process for all n. Thus $\epsilon^{(n)}(t)$ is Gaussian with mean zero and variance

$$\frac{1}{b_n} \int_0^T K^2\left(\frac{t-s}{b_n}\right) ds \le \int_{-1}^1 K^2(u)\,du.$$

Fix $\eta > 0$ and let $k > 1/\delta$. Applying Hölder's inequality on $[0, T]$, Fubini's Theorem, and noting that the $2k$-th moment of $\epsilon^{(n)}(t)$ is uniformly bounded in n and t we get

$$\begin{aligned} E\|\tilde{h} - h^{(n)}\|^{2k} &\le \frac{1}{(nb_n)^k} T^{k-1} \int_0^T E(\epsilon^{(n)}(t))^{2k}\,dt \\ &= O((nb_n)^{-k}) \\ &= O(n^{-k\delta}). \end{aligned}$$

By Chebyshev's inequality

$$P(\|\tilde{h} - h^{(n)}\| > \eta) \le \eta^{-2k} E\|\tilde{h} - h^{(n)}\|^{2k} = O(n^{-k\delta})$$

and since $k\delta > 1$ we have

$$\sum_{n=1}^{\infty} P(\|\tilde{h} - h^{(n)}\| > \eta) < \infty$$

for all $\eta > 0$. The Borel–Cantelli lemma gives $\|\tilde{h} - h^{(n)}\| \xrightarrow{\text{a.s.}} 0$. □

Asymptotic distribution results for estimators of g.

Let $\gamma(t,s)$ be the resolvent kernel for $\Gamma(t,s)$, so the unique solution of (8) is given by

$$g(t) = h(t) + \int_0^t \gamma(t,s)\, h(s)\, ds, \tag{17}$$

see Linz (1985, Theorem 3.3). Note that $h(\cdot)$ may be considered as the output of the linear system

$$\begin{aligned} z'(t) &= [A(t) - C(t)D(t)]z(t) + D(t)g(t), \ z(0) = 0 \\ h(t) &= g(t) - C(t)z(t). \end{aligned}$$

After a trivial manipulation this may be written as a linear system with input $h(\cdot)$ and output $g(\cdot)$:

$$\begin{aligned} z'(t) &= A(t)\, z(t) + D(t)\, h(t), \quad z(0) = 0 \\ g(t) &= C(t)\, z(t) + h(t). \end{aligned}$$

So we may identify the resolvent kernel γ as

$$\gamma(t,s) = C(t)\Psi_A(t,s)D(s), \tag{18}$$

where Ψ_A is the transition function of the system $x'(t) = A(t)x(t)$.

Let the estimator of g corresponding to $\tilde{h}$ be denoted $\tilde{g}$, so that $\tilde{g}$ is the solution of the Volterra integral equation

$$\tilde{h}(t) = \tilde{g}(t) + \int_0^t \Gamma(t,s)\, \tilde{g}(s)\, ds.$$

Now we can write $\tilde{g}$ explicitly as

$$\tilde{g}(t) = \tilde{h}(t) + \int_0^t \gamma(t,s)\, \tilde{h}(s)\, ds. \tag{19}$$

Our next result makes use of (19) to derive the asymptotic distribution of $\tilde{g}$.

THEOREM 4. *Suppose that $nb_n \to \infty$ and $nb_n^3 \to 0$. Then for each $0 < t < T$*

$$(nb_n)^{\frac{1}{2}}(\tilde{g}(t) - g(t)) \xrightarrow{\mathcal{D}} N(0, \kappa^2),$$

where $\kappa^2 = \int_{-1}^1 K^2(u)\, du$.

PROOF: From (17), (19) and the proof of Theorem 3 we have

$$\begin{aligned} (nb_n)^{\frac{1}{2}}(\tilde{g}(t) - g(t)) &= \epsilon^{(n)}(t) + (nb_n)^{\frac{1}{2}}(h^{(n)}(t) - h(t)) \\ &\quad + b_n^{\frac{1}{2}}\eta^{(n)}(t) + (nb_n)^{\frac{1}{2}} \int_0^t \gamma(t,s)(h^{(n)}(s) - h(s))\, ds, \end{aligned} \tag{20}$$

where

$$\eta^{(n)}(t) = \frac{1}{b_n} \int_0^t \int_0^T \gamma(t,s)\, K\left(\frac{s-v}{b_n}\right) dW^{(n)}(v)\, ds.$$

The first term $\epsilon^{(n)}(t)$ on the right hand side of (20) is Gaussian with mean zero and variance

$$\frac{1}{b_n} \int_0^T K^2\left(\frac{t-s}{b_n}\right) ds \to \kappa^2,$$

so that $\epsilon^{(n)}(t) \xrightarrow{\mathcal{D}} N(0,\kappa^2)$. The remaining terms tend to zero in probability. For, using a Fubini theorem for stochastic integrals (see Liptser and Shiryayev, 1977, Theorem 5.15) we have

$$\eta^{(n)}(t) = \frac{1}{b_n} \int_0^T \int_0^t \gamma(t,s)\, K\left(\frac{s-v}{b_n}\right) ds\, dW^{(n)}(v),$$

so that $\eta^{(n)}(t)$ is Gaussian with mean zero and variance

$$\int_0^T \left\{ \frac{1}{b_n} \int_0^t \gamma(t,s)\, K\left(\frac{s-v}{b_n}\right) ds \right\}^2 dv \to \int_0^t \gamma^2(t,s)\, ds.$$

Thus the third term on the right hand side of (20) is of order $O_P(\sqrt{b_n})$. Since h is Lipschitz, the second and fourth terms are of order $O(\sqrt{nb_n^3})$. □

An alternative estimator for g based on the resolvent equation.

An equivalent way of writing equation (17) is

$$g(t) = h(t) + \int_0^t \gamma(t,s)\, dH(s). \tag{21}$$

so an alternative estimator for g is

$$\tilde{g}^a(t) = \tilde{h}(t) + \int_0^t \gamma(t,s)\, d\tilde{H}(s). \tag{22}$$

where $\tilde{h}$ and $\tilde{H}$ are given by (15) and (16). Not surprisingly, $\tilde{g}^a(t)$ has the same asymptotic distribution as $\tilde{g}$.

THEOREM 5. *Suppose that $nb_n \to \infty$ and $nb_n^3 \to 0$. Then for each $0 < t < T$*

$$(nb_n)^{\frac{1}{2}}(\tilde{g}^a(t) - g(t)) \xrightarrow{\mathcal{D}} N(0,\kappa^2).$$

where $\kappa^2 = \int_{-1}^{1} K^2(u)\,du$.

PROOF: From (21), (22) and the proof of Theorem 3 we have

$$(nb_n)^{\frac{1}{2}}(\tilde{g}^a(t) - g(t)) = \epsilon^{(n)}(t) + (nb_n)^{\frac{1}{2}}(h^{(n)}(t) - h(t)) + b_n^{\frac{1}{2}} \int_0^t \gamma(t,s)\,dW^{(n)}(s).$$

By the proof of Theorem 4 the first term $\epsilon^{(n)}(t) \xrightarrow{\mathcal{D}} N(0,\kappa^2)$. The second term is of order $O(\sqrt{nb_n^3})$, as in the proof of Theorem 4. Since the processes $W^{(n)}$, $n \geq 1$ are standard Wiener processes, the last term is of order $O_P(\sqrt{b_n})$. □

4. Trend in the measurement process—periodic case

In this section we consider estimation of g under scheme (II) with $f \equiv 0$ and the functions g, A, B and C assumed to be periodic with period T. We need some preliminary results.

PROPOSITION 6. (Bittanti et al., 1984). *If the pair $(A(\cdot), 1)$ is completely controllable and the pair $(A(\cdot), C(\cdot))$ is completely observable then there exists a unique positive T-periodic solution $\bar{P}$ to the Riccati differential equation (4). Moreover, $\bar{\Psi}$, obtained by replacing P by $\bar{P}$ in the definition of Ψ, is uniformly asymptotically stable, i.e. there exist positive constants K_1 and K_2 such that*

$$|\bar{\Psi}(t,s)| \leq K_1 \exp[-K_2(t-s)], \quad \text{for all } s \leq t.$$

In the scalar case ($p = q = 1$), the pair $(A(\cdot), 1)$ is always completely controllable, and the pair $(A(\cdot), C(\cdot))$ is completely observable under our assumption that $C(\cdot)$ never vanishes on $[0,T]$, see Rubio (1971, Chapter 5). Thus Proposition 6 can be applied directly in that case. Anyway, the hypotheses of Proposition 6 are very natural in the context of linear systems (see Rubio, 1971, Chapter 5).

We shall need the following assumption:

(A) $\int_{-\infty}^{t} \Psi_A(t,s)\,ds < \infty$ and $\int_{-\infty}^{t} \Psi_A^2(t,s)\,ds < \infty$ for all $t \in [0,T]$.

Now introduce the function

$$h_\infty(t) = g(t) + \int_{-\infty}^{t} \bar{\Gamma}(t,s) g(s)\,ds, \tag{23}$$

where $\bar{\Gamma}(t,s) = -C(t)\bar{\Psi}(t,s)C(s)\bar{P}(s)$. Also define $\bar{\gamma}(t,s) = C(t)\Psi_A(t,s)C(s)\bar{P}(s)$. The following lemma shows that there is a useful analogy to the important representation (17) in the periodic case, with the functions h_∞ and $\bar{\gamma}$ playing similar roles to h and γ.

LEMMA 7.
(a) $h_\infty(\cdot)$ *is* T*-periodic.*
(b) *There exists a positive constant* R *such that* $\sup_{t\in[0,T]}|h_i(t)-h_\infty(t)| = O(e^{-Ri})$ as $i\to\infty$.
(c) *If* $\int_{-\infty}^{t}\Psi_A(t,s)\,ds < \infty$, *then*

$$g(t) = h_\infty(t) + \int_{-\infty}^{t}\bar{\gamma}(t,s)h_\infty(s)ds,\ t\in[0,T]. \tag{24}$$

PROOF: From the standard theory of linear O.D.E.'s

$$\bar{\Psi}(t,s) = \exp\left\{\int_s^t [A(u) - C^2(u)\bar{P}(u)]\,du\right\}$$

so that, by the periodicity of A, C and $\bar{P}$, $\bar{\Psi}$ has the property

$$\bar{\Psi}(iT+t, iT+s) = \bar{\Psi}(t,s) \quad \text{for all } s\le t. \tag{25}$$

This property also holds for $\bar{\Gamma}$. Part (a) then follows using the periodicity of g. Next, letting $\bar{h}_i$ be defined by replacing Γ by $\bar{\Gamma}$ in the definition of h_i, we have

$$\begin{aligned}\bar{h}_i(t) &= g(iT+t) + \int_{-iT}^{t}\bar{\Gamma}(iT+t, iT+s)g(iT+s)\,ds\\ &= g(t) + \int_{-iT}^{t}\bar{\Gamma}(t,s)g(s)\,ds\\ &= h_\infty(t) - \int_{-\infty}^{-iT}\bar{\Gamma}(t,s)g(s)\,ds.\end{aligned} \tag{26}$$

Note that $g(s)\bar{\Gamma}(t,s)$ is uniformly asymptotically stable by Proposition 6 and the boundedness of g and C. Thus by (26) and elementary integration, $\sup_{t\in[0,T]}|\bar{h}_i(t) - h_\infty(t)| = O(\epsilon^{-Ri})$ as $i\to\infty$. Here and in what follows, R denotes a generic positive constant which does not depend on T and which may change from use to use. To complete the proof of (b) we need to show that $\sup_{t\in[0,T]}|h_i(t)-\bar{h}_i(t)| = O(\epsilon^{-Ri})$ as $i\to\infty$. Now,

$$\begin{aligned}|h_i(t)-\bar{h}_i(t)| &\le \int_0^{iT+t}|\Gamma(iT+t,s) - \bar{\Gamma}(iT+t,s)|\,ds\\ &\le O(1)\int_0^{iT+t}|\bar{\Psi}(iT+t,s) - \Psi(iT+t,s)|\,ds\\ &\quad + O(1)\int_0^{iT+t}\bar{\Psi}(iT+t,s)|\bar{P}(s)-P(s)|\,ds,\end{aligned} \tag{27}$$

since $P(\cdot)$ is bounded by Roitenberg (1974, p.425). The first term in (27) is bounded above by

$$O(1)\sum_{r=1}^{i}\int_{(r-1)T}^{rT}\bar{\Psi}(iT+t,s)\left|e^{\int_s^{iT+t}C^2(u)[\bar{P}(u)-P(u)]\,du}-1\right|ds$$
$$+O(1)\int_{iT}^{iT+t}\bar{\Psi}(iT+t,s)\left|e^{\int_s^{iT+t}C^2(u)[\bar{P}(u)-P(u)]\,du}-1\right|ds.$$

Use asymptotic stability of $\bar{\Psi}$ to bound the sum of the first $[i/2]$ terms above by $O(e^{-iR})$. We can also bound the sum of the remaining terms by $O(e^{-iR})$ as follows. Writing $P_i(u) = P(iT+u)$ for $0 \le u \le T$, use the rate

$$\sup_{u\in[0,T]}|P_i(u)-\bar{P}(u)| = O(e^{-iR}), \tag{28}$$

given by Roitenberg (1974, Théorème 6, p.431), to obtain for $s \in [(r-1)T, rT]$

$$\begin{aligned}\int_s^{iT+t} C^2(u)|\bar{P}(u)-P(u)|\,du &\le O(1)\sum_{j=r-1}^{i+1}\int_0^T|\bar{P}(u)-P_j(u)|\,du\\ &\le O(1)\sum_{j=r-1}^{i+1}\epsilon^{-Rj} = O(\epsilon^{-Rr}),\end{aligned}$$

uniformly in i. Then, also using the inequality $|\epsilon^x - 1| \le 3|x|$ for $|x| \le 1$, the sum of the "remaining terms" above has the form $\sum_{r=[i/2]+1}^{i+1} O(\epsilon^{-Rr}) = O(\epsilon^{-Ri})$, as required. The last term in (27) is treated in a similar fashion. This proves (b).

Under the hypothesis of part (c), the kernel $\bar{\gamma}$ satisfies $\int_{-\infty}^t \bar{\gamma}(t,s)\,ds < \infty$. Also note that $\bar{\gamma}$ satisfies the property (25). Thus, since g is T-periodic,

$$\begin{aligned}g(t) = g(iT+t) &= h_i(t) + \int_{-iT}^{t}\gamma(iT+t, iT+s)h_i(s)\,ds\\ &= h_i(t) + \int_{-iT}^t \bar{\gamma}(t,s)h_i(s)\,ds + O(1)\int_{-iT}^t \Psi_A(t,s)|P_i(s)-\bar{P}(s)|\,ds\\ &\to h_\infty(t) + \int_{-\infty}^t \bar{\gamma}(t,s)h_\infty(s)\,ds,\end{aligned}$$

as $i \to \infty$, by the dominated convergence theorem, part (b) of the lemma, and (28). This proves (c). □

With the help of Lemma 7 it is now possible to develop results analogous to those of Section 3. For the purposes of illustration we shall discuss kernel estimators.

Define the kernel estimator $\tilde{h}_\infty$ of h_∞ to be the T-periodic function coinciding with $\tilde{h}$ given by (15). Then, in view of (24), it is natural to estimate g by

$$\tilde{g}(t) = \tilde{h}_\infty(t) + \int_{-\infty}^{t} \tilde{\gamma}(t,s)\, \tilde{h}_\infty(s)\, ds, \quad t \in [0,T]. \tag{29}$$

THEOREM 8. *Suppose that (A) holds. Then the entire statement of Theorem 4 carries over to the periodic case, giving the asymptotic distribution of the estimator $\tilde{g}$ defined by (29).*

PROOF: The proof is very similar to the scheme (I) case. Use (24) and (29) to obtain a periodic version of (20):

$$\begin{aligned}(nb_n)^{\frac{1}{2}}(\tilde{g}(t) - g(t)) &= \epsilon^{(n)}(t) + (nb_n)^{\frac{1}{2}}(h_n^*(t) - h_\infty(t)) \\ &\quad + b_n^{\frac{1}{2}}\eta^{(n)}(t) + (nb_n)^{\frac{1}{2}} \int_{-\infty}^{t} \tilde{\gamma}(t,s)(h_n^*(s) - h_\infty(s))\, ds,\end{aligned} \tag{30}$$

where

$$\begin{aligned}\epsilon^{(n)}(t) &= \frac{1}{\sqrt{b_n}} \int_0^T K\left(\frac{t-s}{b_n}\right) dW^{(n)}(s), \\ W^{(n)} &= \sqrt{n}(\tilde{H} - H^{(n)}), \\ H^{(n)}(t) &= \int_0^t \left[\frac{1}{n}\sum_{i=1}^n h_i(s)\right] ds, \quad t \in [0,T]\end{aligned}$$

h_n^* is the T-periodic extension of

$$h_n^*(t) = \frac{1}{b_n}\int_0^T K\left(\frac{t-s}{b_n}\right)\left[\frac{1}{n}\sum_{i=1}^n h_i(s)\right] ds, \quad t \in [0,T]$$

to the whole real line, and

$$\eta^{(n)}(t) = \frac{1}{b_n}\int_{-\infty}^{t}\int_0^T \gamma(t,s)\, K\left(\frac{s-v}{b_n}\right) dW^{(n)}(v)\, ds.$$

It follows from (10′) and (16) that $W^{(n)}$ is a standard Wiener process for all n, so that, as in the proof of Theorem 4, $\epsilon^{(n)}(t) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\kappa^2)$. By Lemma 7 (b)

$$\sup_{s\in[0,T]} \left|\frac{1}{n}\sum_{i=1}^n h_i(s) - h_\infty(s)\right| = O(n^{-1})$$

so that, since h_∞ is Lipschitz, the second term on the right hand side of (30) is of order

$$(nb_n)^{\frac{1}{2}}(h_n^*(t) - h_\infty(t)) = O\Big(\frac{b_n}{n}\Big)^{\frac{1}{2}} + O(\sqrt{nb_n^3}).$$

Using Condition (A) it can be shown that the last term on the right hand side of (30) is of the same order. Using Condition (A) again, the third term can be shown to be of order $O_P(\sqrt{b_n})$, as in the proof of Theorem 4. □

5. Trend in the state process

Throughout this section it is assumed that the trend in the measurement process is zero. We shall introduce an estimator $\tilde{f}$ of the trend f in the state process such that $BIAS(\tilde{f}, 0, t) \to 0$ uniformly in t as $n \to \infty$ a.s.. In fact $\tilde{f}$ is shown to be strongly L^2-consistent. We shall only consider the case of observation scheme (I) since our results can be extended easily to scheme (II) along the lines that we extended our results on estimation of g in Section 4.

To estimate f we need to consider (7), which is a linear Volterra integral equation of the first kind. The usual way to deal with such equations is to convert them into Volterra equations of the second kind by differentiation, see Linz (1985, p.67). In fact, using this technique, we may solve (7) explicitly for f. Since $C(t)$ does not vanish on $[0, T]$, we obtain

$$f(t) = \frac{h'(t)}{C(t)} + F(t)h(t) \tag{31}$$

where

$$F(t) = D(t) - \frac{C'(t)}{C^2(t)} - \frac{A(t)}{C(t)}.$$

Thus the problem of estimating f is similar to the problem of estimating g, except that now we need to estimate h' as well as h. We shall only consider kernel estimators of h', although trigonometric series sieve estimators (see Ibragimov and Khasminski, 1980) could equally well be used.

Let K be a kernel function, as in Section 3, but in addition assume that K is differentiable. Let c_n be a bandwidth parameter, different from b_n. Define

$$\tilde{h}'(t) = \frac{1}{c_n^2}\int_0^T K'\Big(\frac{t-s}{c_n}\Big)d\tilde{H}(s). \tag{32}$$

The following result, stated without proof, is similar to Theorem 3.

THEOREM 9. *Suppose that* $c_n \to 0$ *and* $c_n n^{\frac{1}{3}-\delta} \to \infty$ *where* $0 < \delta < \frac{1}{3}$. *Then the kernel estimator* $\tilde{h}'$ *given by (32) is a strongly* L^2*-consistent estimator of* h'.

In view of this result and (31) it is reasonable to estimate f by

$$\tilde{f}(t) = \frac{\tilde{h}'(t)}{C(t)} + F(t)\tilde{h}(t), \tag{33}$$

where $\tilde{h}$, given by (15), is the kernel estimator of h. Under the joint conditions of Theorems 3 and 9 we see that $\tilde{f}$ is a strongly L^2-consistent estimator of f. Finally, we give an asymptotic distribution result for $\tilde{f}$.

THEOREM 10. *Suppose that* $nb_n \to \infty$, $nb_n^3 \to 0$, $nc_n^3 \to \infty$, $nc_n^5 \to 0$ *and* $c_n = o(b_n^{\frac{1}{3}})$. *Then for each* $0 < t < T$

$$(nc_n^3)^{\frac{1}{2}}(\tilde{f}(t) - f(t)) \xrightarrow{\mathcal{D}} N(0, \sigma^2(t)),$$

where

$$\sigma^2(t) = \frac{\int_{-1}^{1} K'(u)^2 \, du}{(C(t))^2}.$$

PROOF: Directly from (31) and (33)

$$\begin{aligned}(nc_n^3)^{\frac{1}{2}}(\tilde{f}(t) - f(t)) = (C(t))^{-1}(nc_n^3)^{\frac{1}{2}}(\tilde{h}'(t) - h'(t)) \\ + F(t)\Big(\frac{c_n^3}{b_n}\Big)^{\frac{1}{2}}(nb_n)^{\frac{1}{2}}(\tilde{h}(t) - h(t)).\end{aligned}$$

It can be shown, using a similar approach to the proof of Theorem 4, that the first term on the right hand side tends in distribution to $N(0, \sigma^2(t))$. Also from the proof of Theorem 4, and using the condition $c_n = o(b_n^{\frac{1}{3}})$, the second term on the right hand side is seen to be of order $o_P(1)$. □

6. The multivariate case

In the general case in which the state and measurement processes are p and q-dimensional, our results are modified in obvious ways to take into account the fact that A, B, C etc. are matrices. The innovations process ν is now a q-dimensional Wiener process and (4) is replaced by the matrix Riccati equation

$$P'(t) = A(t)P(t) + P(t)A(t)^T - P(t)C(t)^T C(t)P(t) + I,$$

with initial condition $P(0)$ = covariance matrix of $X(0)$. Here I is the $p \times p$ identity matrix, and "^{T}" denotes "transcript." The Kalman gain is now given by $D(t) = P(t)C(t)^T$.

In Section 3 the q-dimensional version of the orthogonal series sieve estimator $\hat{h}$ is defined (using the same sieve for each component of g) by

$$\hat{h}_k(t) = \sum_{r=1}^{d_n} \hat{h}_{kr}\, \psi_r(t),$$

$k = 1, \ldots, q$, where

$$\hat{h}_{kr} = \frac{1}{n} \sum_{i=1}^{n} \int_0^T \psi_r(t)(dY_{ik}(t) - U_{ik}(t)\, dt).$$

The kernel estimator $\tilde{h}$ is defined (using the same kernel function and bandwidth for each component of g) by the q-dimensional version of (15). The estimators $\hat{g}$ and $\tilde{g}$ are defined as before. Theorems 1–5 extend with the modification that the limiting distribution in Theorems 4 and 5 is $N(0, \kappa^2 I)$. In the proofs of these results, $W^{(n)}$ becomes a q-dimensional Wiener process.

For the results of Section 4 to hold, the additional assumptions that $(A(\cdot), I)$ is completely controllable and $(A(\cdot), C(\cdot))$ is completely observable are needed. Condition (A) becomes

(A) $\int_{-\infty}^{t} \|\Psi_A(t,s)\|\, ds < \infty$ and $\int_{-\infty}^{t} \|\Psi_A(t,s)\|^2\, ds < \infty$ for all $t \in [0, T]$.

Here $\|\cdot\|$ denotes operator norm. There is essentially no change in the proofs, with the results of Bittanti et al. (1984) and Roitenberg (1974) being applied in the same way as before.

The results of Section 5 extend under the condition that for each $t \in [0, T]$ the matrix $C(t)$ has a left inverse $C^{-1}(t)$. This will be the case if $p \le q$ and $C(t)$ has column rank p for each $t \in [0, T]$. Then (31) becomes

$$f(t) = C^{-1}(t)h'(t) + F(t)h(t),$$

where

$$F(t) = D(t)C(t)C^{-1}(t) - C^{-1}(t)C'(t)C^{-1}(t) - A(t)C^{-1}(t),$$

showing that f is identifiable. Note that f is not identifiable if $p > q$. The limiting distribution in Theorem 10 becomes $N(0, \Sigma(t))$, where

$$\Sigma(t) = C^{-1}(t)C^{-1}(t)^T \int_{-1}^{1} K'(u)^2\, du.$$

7. Directions for further work

The techniques and results developed in this article are by no means exhaustive. We are aware of many important questions concerning the problem of nonparametric inference for linear systems in continuous time for which we have no answer at this stage. We conclude by listing some of these questions, the first two of which were mentioned by a referee.

(1) Is it possible to weaken the assumption that A, B and C be known? How far would the analysis go say, if B was unknown? (This would require an assumption of sufficient variability in the deterministic inputs u_i, $i \geq 1$ to avoid an identifiability problem.) In the same line, how robust are the estimators of f and g to the specification of A, B and C?

(2) Can anything be said about the optimal choice of the bandwidth in the kernel estimators? In the cases of density estimation and nonparametric curve estimation there are various techniques for automatically selecting the bandwidth. It ought to be possible to develop such methods of 'cross-validation' here.

(3) Can a test for detecting the *presence of a trend* (e.g. a test of $g \neq 0$) be developed? More generally, it is of interest to test of whether the trend is of some specified form. As in the case of goodness-of-fit testing for distribution functions, this might be done by deriving a functional central limit theorem for an estimator of the cumulative trend function $G(\cdot) = \int_0^{\cdot} g(s)\, ds$.

References

Aihara, S. and Bagchi, A. (1989). Infinite dimensional parameter identification for stochastic parabolic systems. *Statist. Prob. Letters*, to appear.

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1988). Censoring, truncation and filtering in statistical models based on counting processes. *Contemporary Math.* **80** 19–59.

Arjas, E. and Haara, P. (1984). A marked point process approach to censored failure data with complicated covariates. *Scand. J. Statist.* **11** 193–209.

Bagchi, A. and Borkar, V. (1984). Parameter identification on infinite dimensional linear systems. *Stochastics* **12** 201–213.

Balakrishnan, A. V. (1973). *Stochastic Differential Systems*. Lecture notes in economics and mathematical systems. Springer–Verlag, Berlin.

Beder, J. H. (1987). A sieve estimator for the mean of a Gaussian process. *Ann. Statist.* **15** 59–78.

Bittanti, S., Colaneri, P. and Guardabassi, G. (1984). Periodic solutions of periodic Riccati equations. *IEEE Trans. Automatic Control* AC-**29** 665–667.

Bittanti, S. and Guardabassi, G. (1986). Optimal periodic control and periodic systems analysis: An overview. In *Proc. 25th Conf. Decision Control*, IFAC, Athens, Greece, 1986, pp. 1417–1423.

Davis, M. H. A. (1977). *Linear Estimation and Control*. Chapman and Hall, London.

Davis, M. H. A. and Vinter, R. B. (1985). *Stochastic Modelling and Control*. Chapman and Hall, London.

Geman, S. and Hwang, C. R. (1982). Nonparametric maximum likelihood estimation by the method of sieves. *Ann. Statist.* **10** 401–414.

Greenwood, P. E. (1988). Partially specified semimartingale experiments. *Contemporary Math.* **80** 1–17.

Grenander, U. (1981). *Abstract Inference*. Wiley, New York.

Ibragimov, I. A. and Khasminski, R. Z. (1981). *Statistical Estimation*. Springer-Verlag, New York.

Ibragimov, I. A. and Khasminski, R. Z. (1980). Estimates of the signal, its derivatives, and point of maximum for Gaussian distributions. *Theory Prob. Appl.* **25** 703–719.

Jazwinski, A. H. (1970). *Stochastic processes and filtering theory*. Academic Press, New York.

Kallianpur, G. (1980). *Stochastic Filtering Theory*. Springer-Verlag, New York.

Kalman, R. E. and Bucy, R. S. (1961). New results in linear filtering and prediction theory. *Trans. ASME Ser. D, J. of Basic Eng.* **83** 95–108.

Kumar, P. R. and Varaiya, P. (1986). *Stochastic Systems: Estimation, Identification and Adaptive Control*. Prentice Hall, Englewood Cliffs, N.J.

Leskow, J. (1989). Maximum likelihood estimator for almost periodic stochastic process models. Tech. Report, Center for Stochastic Processes, Dept. of Statistics, University of North Carolina at Chapel Hill.

Linz, P. (1985). *Analytical and Nummerical Methods for Volterra Equations*. SIAM, Philadelphia.

Liptser, R. S. and Shiryayev, A. N. (1977). *Statistics of Random Processes I: General Theory*. Springer-Verlag, New York.

Liptser, R. S. and Shiryayev, A. N. (1978). *Statistics of Random Processes II: Applications*. Springer-Verlag, New York.

McKeague, I. W. (1986). Estimation for a semimartingale regression model using the method of sieves. *Ann. Statist.* **14** 579–589.

McKeague, I. W. (1988). A counting process approach to the regression analysis of grouped survival data. *Stoch. Process. Appl.* **28** 221–239.

Nguyen, P. A. and Pham, T. D. (1982). Identification of nonstationary diffusion model by the method of sieves. *SIAM J. Control Optim.* **20** 603–611.

Pons O. and de Turckheim, E. (1988). Cox's periodic regression model. *Ann. Statist.* **16** 678–693.

Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11** 453–466.

Roitenberg, I. (1974). *Théorie du Controle Automatique.* Mir, Moscow.

Rubio, J. E. (1971). *The Theory of Linear Systems.* Academic Press, New York.

Tugnait, J. K. (1980). Identification and model approximation for continuous-time systems on finite parameter sets. *IEEE Trans. Automatic Control* AC-25 1202–1206.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER ARO 23699.22-MA	2. GOVT ACCESSION NO. N/A	3. RECIPIENT'S CATALOG NUMBER N/A
4. TITLE (and Subtitle) Nonparametric estimation of trends in linear stochastic systems		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) McKeague, Ian W. and Tofoni, Tiziano		8. CONTRACT OR GRANT NUMBER(s) DAAL03-86-K-0094
9. PERFORMING ORGANIZATION NAME AND ADDRESS Florida State University Department of Statistics Tallahassee, FL 32306-3033		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE September 1989
		13. NUMBER OF PAGES 21
		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

N/A

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Kalman-Bucy, linear periodic systems, nonparametric trend estimation, mathod of sieves, kernel estimators, Volterra integral equations.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Techniques for the estimation of unknown additive trends present in the state and measurement processes of a Kalman–Bucy linear system are introduced. We obtain asymptotic results describing the performance of the estimators under i.i.d. and periodic observation schemes. The observed process is given by $dY(t) = g(t)\,dt + dZ(t)$, where Z is the measurement process and g is an unknown trend function, and there is an additive trend f present in the state process X. These two cases need to be treated separately in order

continued . . .

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract continued . . .

to ensure identifiability. The problem is to estimate f and g, and remove them from the measurement process. Trend removal involves replacing f and g in the Kalman filter $\hat{X}(t) = E(X(t)|\mathcal{F}_t^Y)$—based on observation of Y—by appropriate estimates. We show that this can be done under the following observation schemes: (I) n i.i.d. replicates of Y over a fixed interval $[0,T]$, (II) observation of a single trajectory of Y over a long interval $[0,nT]$, where f, g and the functions defining the linear system are periodic with period T.